Mike John Page

Data Scientist & R Specialist

807 Jefferson Place · Manchester · M4 4BL, UK

✉ m.page.j@gmail.com ⬡ mikejohnpage 🌐 mikejohnpage.com

Personal Statement

A published data scientist and R specialist that is highly motivated, analytical, and revels on intellectual challenges. Currently a Junior Data Scientist at DataKolektiv that has developed the newsrivr R package and assisted in the creation of a new information retrieval schema. Has a proven track-record working as a freelancer helping start-ups receive funding. Excellent communicator who thrives in creating a story from data and works diligently in interdisciplinary teams to deliver timely results.

Relevant Data Skills

• *Technologies:* R, Python, MySQL, PostgresSQL, Bash, Git, Markdown, SPSS Statistics, Excel, G-Power, Linux
• *Skills:* remote servers, headless browsers, data mining, web scraping, API clients, data wrangling, exploratory analysis, machine learning and statistics, data visualisation
• *Data libraries:* tidyverse, NumPy, pandas, httr, jsonlite, xml2, widyr, blogdown, topic-models

Completed Projects

Perfectionism in the Public Domain: A Natural Language Processing Approach: built an API client in R using the *httr* package that gathered news articles from newsriver.io. Using a tidy data approach, I then performed sentiment analysis and topic models using latent Dirichlet allocation.

Web Scraping the Weather: used Python to build a web scraper that collects the current weather forecast for a given location from Wunderground. As the weather information was rendered in JavaScript, Selenium was used in combination with Morzilla's gecko-driver. The retrieved content was then parsed using Beautiful Soup. The script was then set up as a bash alias to be executed daily, as described in my blog post here.

Multilevel Modelling in SPSS: wrote a script in SPSS syntax that executed multilevel (hierarchical linear) models on nested physiological and psychological data.

GGDOT Hacknight: I ran Box-Cox transformations and regression in R to analyse the CO_2 emissions of food. I created new features in the data set and found a novel and unexpected relationship in the data. The finding is now being further validated by the GGDOT group before being shared with the Public health community.

Freelance Project: Drawing Insight from Simple Data Exploration and Visualisation: used choropleth maps, dodged bar charts and other simple visualisation techniques (alongside data exploration) to inform a new start-up of competitors, industry trends, and

recommended areas for investment. The data and figures were used successfully by the company to secure independent funding.

Ongoing Projects

newsrivr: I am writing a R client for Newsriver's REST API. The package contains functions to secure credentials, retrieve JSON files from the API with specified query parameters, and clean the retrieved data following a set of sensible defaults.

Opt Out: I am a collaborator on Opt Out, a browser extension that helps people protect themselves against sexual harassment and sexist hate speech on Facebook and Twitter. My key roles include reviewing the literature, deploying scripts to AWS to retrieve data, and evaluating and tuning binary classifier models.

Project Contributions

newsanchor: I rewrote the `set_api_key()` function to be more secure by using the askpass package to prompt users for their credentials. This follows best security practices as it prevents personal credentials from being stored in the .Rhistory file which can be accidentally leaked.

Education

Springboard: *Introduction to Data Science*   2018

- A 100-hour online data science workshop
- R, Data Wrangling, Analytics, and Data Storytelling
- Capstone project reviewed by an independent industry expert

DataCamp   2018

- Completed thirteen courses
- Python, R, Programming, Importing and Cleaning Data, Probability and Statistics

York St John University: *MSc by Research in Psychoendocrinology*   2016-2017

- Awarded £2500 Nestlé research grant
- Won poster presentation prize in recognition of excellence in the presentation of doctoral research at the 2016 BSPID conference
- Mastered SPSS, Excel, G-Power and complex hierarchical linear models

York St John University: *Sport Science Performance Conditioning BSc*   2010-2014

- 1st Class Honours
- Dissertation prize, highest course grade prize

Employment

Freelance Data Scientist   current

- deliver a range of work from web scraping to to machine learning applications

DataKolektiv - Junior Data Scientist   2018-current

- developed the open source newsrivr R package
- assisted in the development of an information retrieval schema that enhances the annotation of text data via additional meta features
- completed tasks that improved skills in data scraping and headless Linux environments

York St John University

Lecturer and Module Director: Sport and Exercise Biomechanics   2017-2018

- Curated and delivered module including lectures, labs, and assessments
- Increased mean student grade from previous years
- Collaborated with laboratory technician to organise materials for labs

Laboratory Technician: Biomechanics laboratory   2016

- Prepared equipment for teaching and research activities
- Liaised with industry partners to resolve technical issues
- Contributed to teaching and research environment

Research Assistant: Biomechanics   2014

- Analysed 3D modelling data using a Qualisys motion capture system and Visual3D
- Performed in depth literature reviews
- Managed and analysed large, complex data sets

Publications

Page, M.J., Hill, A.P., Kavanagh, O., & Jones, S. (2018). Multidimensional perfectionism and cortisol response in non-clinical populations: A systematic review and evaluation. *Journal of Personality and Individual Differences, 124,* 16-24.

Page, M.J., Hill, A.P., Kavanagh, O., & Jones, S. (2018). Cortisol Awakening Response as an Indicator of Pre-Competition Stress in Elite and Non-Elite Athletes. *Publication date TBC.*

Additional Languages

Italian: *conversational* | Spanish: *A1*

Other

Charity: cycled over 6800 Km unsupported across Europe to raise over £7000 for various charities.

Interests: Mountain biking, rock climbing, surfing, travelling, playing guitar.

References available upon request.